# Bike Prices: Feature Engineering and Exploratory Data Analysis (EDA)

## Project Description

This project aims to predict used bike prices using machine learning techniques. By analyzing historical data on bike sales, we explore various features such as model year, kilometers driven, power, and more to understand their impact on pricing. The objective is to develop a predictive model that can accurately forecast bike prices based on these features, aiding potential buyers and sellers in the market.

## Data Source

The dataset used in this analysis is derived from a collection of used bike sales records. It includes details such as model name, year, kilometers driven, ownership, location, mileage, power, and the sale price. This data is compiled to form a comprehensive overview of the used bike market, facilitating the exploration of trends and patterns in bike pricing.

## Methodology

The methodology involves several key steps: data preprocessing to handle missing values and outliers, exploratory data analysis (EDA) to identify trends and relationships, feature engineering to enhance the predictive capability of the model, and linear regression modeling to predict bike prices. Both univariate and multivariate analyses are conducted to explore individual features and their combined effects on bike prices.

## Results and Analysis

Preliminary results indicate strong correlations between certain features and the bike price. The model year and kilometers driven appear to significantly influence the price, with newer, less-used bikes fetching higher prices. Power and owner status also play crucial roles, with higher power bikes and those with fewer previous owners commanding premium prices. Visualizations from the EDA provide clear insights into these trends.

## Conclusions and Recommendations

The analysis confirms that certain features are strong predictors of bike prices, which can be effectively used in a predictive model. We recommend further refinement of the model with additional features such as bike brand and model specifics, and possibly external economic factors. For stakeholders in the used bike market, these insights can aid in pricing strategies and inventory management. Further research could explore more complex models such as ensemble methods or neural networks for potentially higher accuracy.